



ANALYSIS OF BUSINESS INTELLIGENCE TOOLS

Madhura Gajendragadkar¹ | Rohit Bapat¹ | Shailesh P Bendale¹ | Rohit Chougule¹ | Harshvardhan Kadam¹

¹ Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune, India.

ABSTRACT

In recent times, technology applications in Business Intelligence (BI) have been developed rapidly. BI is considered to be one of the hottest emerging technologies. BI consists of a broad category of applications and technologies for gathering and storing data to analyze and help enterprises make smarter business decisions and strategies for designing and planning. As opposed to this in the past, the Business Intelligence market was strictly dominated by closed source and commercial tools, but in recent times, some open source solutions have been made available. The BI market is exciting, continually innovating and growing to meet the ever-expanding requirements of businesses of all sizes and industries. This represents a vast competitive advantage; however, the choice of an appropriate open source BI suite is a challenge. An understanding of BI tool categories is needed so as to match your analytical need with the appropriate tools. The current study evaluates and compares the latest versions of the two main Open Source Business Intelligence suites: Jaspersoft and Pentaho.

KEYWORDS: Business Intelligence, Pentaho, Jaspersoft, data integration.

I. INTRODUCTION

Business intelligence (BI) can be described as "a set of techniques and tools for the acquisition and transformation of raw data into meaningful and useful information for business analysis purposes"[1]. The goal of BI is to allow for easy interpretation of huge volumes of data and to identify relevant parts of this data which can be used for improving business strategies so as to generate more profit. This can provide businesses with a competitive market advantage and long-term stability[2].

BI applications include the following:

Reporting: for representing data in a readable format.

Online analytical processing: for data analysis in warehouses.

Analytics: for analyzing trends and patterns.

Data mining: for digging up useful and relevant data from a huge amount of data.

Business performance management: to increase future profits.

Benchmarking: to set a minimum value.

Text mining: to search for specific keywords.

BI tools empower organizations in gaining insight into new markets. [3] BI solutions based on big data use various filtering tools to gather only the required data, and enable a continuous analysis on streaming data. [4]

The use of big data and business intelligence has already transformed business decision-making in companies such as Amazon.com and Walmart[5][6][7].

Consider this example:

A store has multiple branches across many cities. It sells the same products everywhere for the same rate. During a monthly analysis, it was found that the sales of a particular product were higher in a certain branch as compared to all other branches. The store can then divert the stock of that product to this branch where more sales and therefore more profits, are being made. This is an application of business intelligence.

Another common example is when stores stock certain items together. For instance, dairy products and milk are kept close by in most stores. This is because customers are likely to purchase these items together frequently. Because they are placed side by side, a customer is potentially attracted to making both purchases as opposed to if he had to go all the way across to the other end of the store. Another strategy is to deliberately place frequently bought items some distance away so as to make the customer go through other items like cookies and chips on the way, which he may then be tempted to buy.

II. RELATED TOOLS:

A. Pentaho:

The Pentaho suite is currently the most used and reputed BI tool in the market [8]. Pentaho was founded in 2004 by a group of executives with extensive experience in BI coming from companies like Business Objects, Cognos, Hyperion, JBoss,

Oracle, Red Hat and SAS. The Pentaho suite consists of two versions, the Enterprise and Community versions. The Enterprise version is a paid distribution that contains all features with user-friendly interfaces and is easy to understand for the end users. The Community version is open source, free, and provides virtually the same modules, some of which do not have user friendly interfaces, and has a greater difficulty in understanding and implementation. Its development is done in Java, and can be run from a Java Virtual Machine. The Community version is composed of modules Pentaho Data Integration, Pentaho Analysis Services, Pentaho Dashboards, Pentaho Data Mining and Pentaho Reporting. The Pentaho Data Integration, also known as Kettle, allows performing the ETL data easily and intuitively. It supports a large library of mapping objects to support multiple data sources, and allows data storage for data warehouses of various dimensions, other data files or databases. The transformation enables data cleansing through rules and migration of data between applications. The Pentaho Reporting consists of two tools, a reporting, also known as JFreeReport, and another for generating metadata, which allows the creation of ad-hoc Web reports. It supports various data sources including relational, OLAP and XML. It also allows you to design reports supported by wizard, multi-lingual, interactive display filters, and export to the most common formats. Performance indicators are guaranteed by Pentaho Dashboards, which allows you to create control panels, and gather in the same window the main indicators of a department or the whole company. The metrics are available in an intuitive way, and allows integration with Pentaho Reporting and Analysis. It also provides continuous monitoring with problems notification alerts. The Pentaho Analysis Services, also known as Mondrian is an OLAP engine, based on a ROLAP architecture, which can be used with major database management systems, and has features such as metadata layer, MDX, cache memory, aggregate tables, etc. A fully supported Web environment allows you to create reports with drag & drop support, display graphics, export multi-dimensional information and visualization with selection of metrics and attributes to be analyzed. Also known as Weka, Pentaho Data Mining is a set of tools for data mining, whose set of classification rules, regression, association and clustering algorithms help to better understand the business and improve future performance. The Community Edition is distributed under the GNU General Public License version 2.0 licenses (GPLv2), GNU Lesser General Public License version 2.0 (LGPLv2) and Mozilla Public License 1.1 (MPL 1.1). The latest version of this suite is the 4.8, released in November 2012.

B. Jaspersoft:

Jaspersoft is known for providing a form of self-service tailored to the individual needs of companies [8]. The Jaspersoft suite was created in 2006, after several years the company has created various tools individually. Jaspersoft provides the most flexible BI, which is economical and widely deployed in the world. The Jaspersoft website states that more than 14 million copies of open source software have been downloaded in the world, with 175,000 production deployments and over 14,000 customers in 100 countries. They also claim that the suite is updated frequently by a development community of more than 225,000 registered members. This community has the distribution of open source, free, and commercial distribution spread over three editions (Express, Professional and Enterprise). The distribution community is very limited as compared to commercial distributions and is distributed under a GNU GPL. It is composed of Jaspersoft ETL, JasperReports Server modules, Jaspersoft Studio, JasperReports Library, and iReport Designer.

JasperReports Server is a standalone reporting server that can be embedded in any Java application. It provides reports and analysis that can be incorporated into Web applications or mobile applications. It also provides real-time or scheduled reports to the web, mobile, printer or e-mail in a variety of formats. It is optimized to share, protect and centrally manage the reports and analysis. Among the various features the ones that stand out are: formatting and interactively viewing reports, centralized and secure repository, generation, scheduling and distribution of reports and customizable interface. According to Jaspersoft, JasperReports Library is the most popular open source tool for creating reports. It is entirely programmed in Java and is able to use data from any source and produce documents which can be viewed, printed or exported in a variety of formats including HTML, PDF, Excel, OpenOffice and MSWord. Data integration (ETL - Extract, Transform, and Load) is supported by Jaspersoft ETL. This allows you to extract data from multiple sources, transform the data based on defined business rules and loads them into a data warehouse or data mart for analysis and reporting. Among the features include the graphical desktop environment, more than 500 connections to components and version control work.

The Jaspersoft Studio is a report design environment based on Eclipse for JasperReports and JasperReports Server. It lets you create reports from any data source, formatted for viewing on screen or print format, exporting to a wide range of formats. Among the features include the graphical desktop environment, the reporting models supported by themes, integration with JasperReports Server, sophisticated layouts with graphics, images, cross-sub-reports and tables, access data via JDBC, TableModels, JavaBeans, XML, Hibernate, CSV and custom backgrounds and publish the reports in PDF, CSV, RTF, XLS, XML, HTML, DOCX, text files, or OpenOffice. Based on NetBeans, iReport Designer, Jaspersoft Studio is a tool to image, with essentially the same features. The latest version of Jaspersoft Studio is 5.0, the last update of December 2012.

III. CONCLUSION & FUTURE WORK

After considering multiple factors for comparison between Jaspersoft and Pentaho, it is clear that each tool has its own pros and cons which will impact different business applications differently. The choice of the tool should be based on the goal of the analytics as each tool has advantages in different aspects. As concluded from the current study, Pentaho has the greatest potential for use in a corporate environment.

As future work, we intend to continue this study assessing the suites and other technical usability. We intend to carry out the implementation in a real-world environment, for market basket analysis and to plan improved promotion strategies for coupons and discounts. Pentaho is more suited to our proposed work.

IV. REFERENCES

1. Turner, Dawn M. "What is Venture Management?". www.VentureSkies.com. VentureSkies. Retrieved 24 February 2016.
2. Rud, Olivia (2009). *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy*. Hoboken, N.J: Wiley & Sons. ISBN 978-0-470-39240-9.
3. Chugh, R & Grandhi, S 2013, 'Why Business Intelligence? Significance of Business Intelligence tools and integrating BI governance with corporate governance', *International Journal of E-Entrepreneurship and Innovation*, vol. 4, no.2, pp. 1-14.
4. Yesudas, Menon, Ramamurthy, 'Intelligent Operational Dashboards for smarter commerce using big data', *IBM J. Res. and Dev.*, Volume 58, Paper 13, September/November 2014.
5. V. M. Scho'nberger and C. Kenneth, *Big Data: A Revolution That Will Transform How We Live, Work, Think*. Boston, MA, USA: Houghton Mifflin, 2013.
6. M. Sean, *How Companies Like Amazon Use Big Data To Make You Love Them*. [Online]. Available: <http://www.fastcodesign.com/1669551/how-companies-like-amazon-use-big-data-to-makeyou-love-them>
7. K. Noelle, *Now Trending: Big Data at Walmart.com*. [Online]. Available: <http://blogs.wsj.com/cfo/2013/11/22/now-trending-bigdata-at-walmart-com>
8. Marinheiro A. and Bernardino J., "Analysis of Open Source Business Intelligence Suites", 2015.